Online Appendix: Coronavirus Infections and Deaths by Poverty Status: The Effects of Social Distancing

A Fitting Curves Using the Leave-One-Out Method

This section describes the curve fitting method we use to produce the smooth curves in Figures 4–7. We start with 20 points that are calculated using the non-parametric method described in Section 3.[31] We then employ a local linear estimation that results in a non-parametric fit that incorporates these 20 points. However, the fit depends on a smoothing parameter. If the smoothing parameter is very high, the curve becomes the the best fit line of an OLS estimate. If the smoothing parameter is low, noise increases and the lines starts to move through every point. Fitting a smooth curve through the 20 points becomes a trade-off between bias (using high value and producing a very smooth curve) and noise (using a low value). We use a procedure that minimizes the residual mean squared error (RMSE) from a prediction resulting from leaving one of the 20 points out when estimating a local regression.

This leave-one-out cross-validation method minimizes the RMSE but is robust to the possibility of in-sample over fitting. This method works as follows. In the case of 20 points, we first start with the starting value of the smoothing parameter α_1. We use the last 19 points (excluding the first point) and estimate the local linear model. Then we use this estimation to predict the value of the first point we left out. The difference between the first (actual point) and the first (predicted point) contributes to the MSE. We perform similar estimations by excluding each point of the 20 points and using the resulting 19 to perform local linear regression. We then perform similar out-of-sample predictions and use the excluded point to calculate the RMSE. The RMSE for the first value of the starting smoothing parameter α_1 is RMSE(α_1) $= \sum_{i=1}^{20} \frac{(x_i - \hat{x}_{\alpha_1,i})^2}{20}$ where x_i is the actual point observation and $\hat{x}_{\alpha_1,i}$ is the prediction of point i based on local regressions using smoothing parameter α_1. We next repeat this for tightly packed values of the smoothing parameter $\alpha \in [\underline{\alpha}, \bar{\alpha}]$, which gives a series of RMSE$(\alpha)_\alpha$. We then choose the minimum RMSE and its associated smoothing parameter $\hat{\alpha}$.

[31] This method ranks all US counties according to percentage of individuals living under the poverty level and then forms 20 county-groups of roughly equal population size. Each group is an observation in Figures 4–7.

B Results Based on 25 County Groupings

We replicate our descriptive findings based on 20 county-groups, originally presented in Figures 4 and 5, using 25 county-groups instead. The results, shown in Figures B.1 and B.2, are very similar to the patterns shown in Figures 4 and 5, suggesting that the descriptive patterns are robust to alternative groupings of counties.

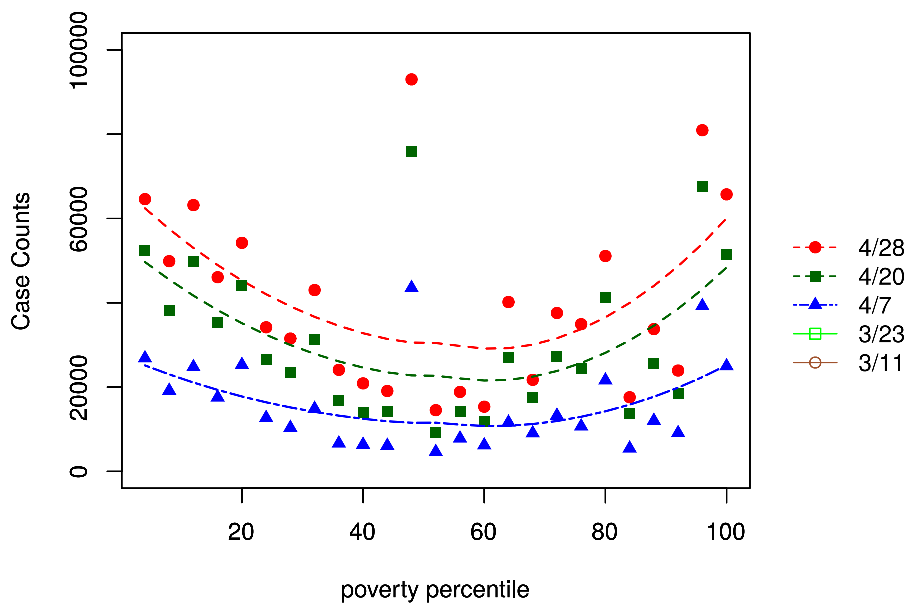

Figure B.1: **SARS-CoV-2 Confirmed Infections by Poverty Percentile Based on 25 County Groups**

Notes: The figure replicates Figure 4 except that we form 25 county-group bins based on the poverty rate. The source of data is USAFacts, as of April 28, 2020. We report the number of cumulative (infection) cases for March 11, March 23, April 7, April 20, and April 28 in 2020 for 25 county-groups ranked by poverty rate percentile. The curves are fitted using a smoothing method based on local linear regressions as described in Appendix A.

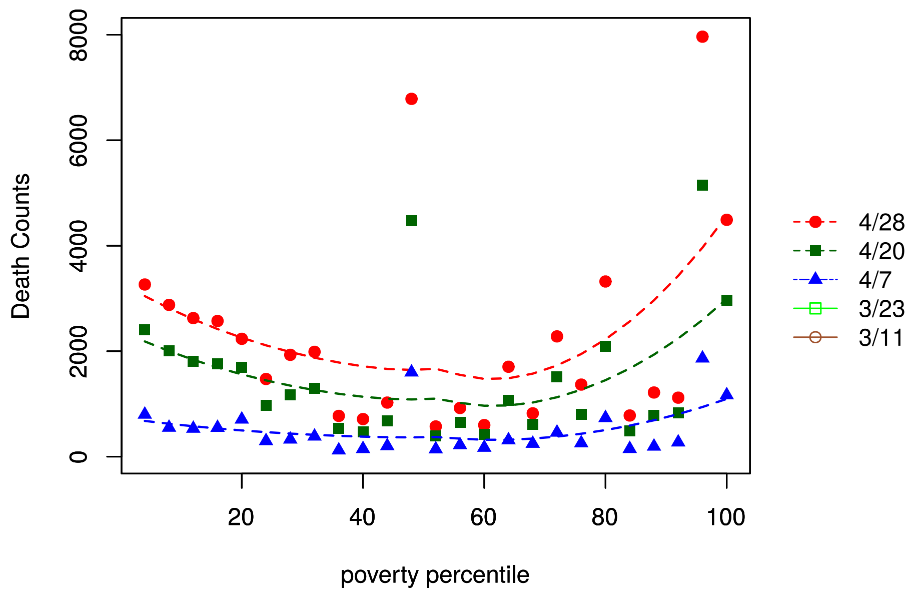

Figure B.2: **SARS-CoV-2 Confirmed Deaths by Poverty Percentile Based on 25 County Groups**

Notes: The figure replicates Figure 5 except that we form 25 county-group bins based on the poverty rate. The source of data is USAFacts, as of April 28, 2020. We report the number of cumulative deaths due to COVID-19 for March 11, March 23, April 7, April 20, and April 28 in 2020 for 25 county-groups ranked by poverty rate percentile. The curves are fitted using a smoothing method based on local linear regressions as described in Appendix A.

C Removing County Observations from the State of New York

From the descriptive analysis in Section 2.3 it is obvious that the median county-group is an outlier, not following the U-shape. This specific county-group contains Queens in the state of New York. Queens reported the highest infection and death rates in the early days of the pandemic.[32] Given that this county was hit especially hard, we next repeat our descriptive analysis without any county observations from the state of New York. In addition, we re-estimate specification 1, 2, and 3 with this reduced sample in order to check whether our results are driven by the dramatic surge in COVID-19 cases in New York during the onset of the pandemic.

The descriptive analysis, shown in Figures C.1 and C.2, are similar to the main analysis in Figures 4 and 5 of Section 2.3 in the main paper. Similarly, the estimation results concerning the interaction coefficients presented in Figure C.3 are very similar to the main results in Figure 9, again suggesting that our findings are robust with respect to observations from the state of New York.

[32] See for instance a Time article from April 5, 2020: https://time.com/5815820/data-new-york-low-income-neighborhoods-coronavirus/

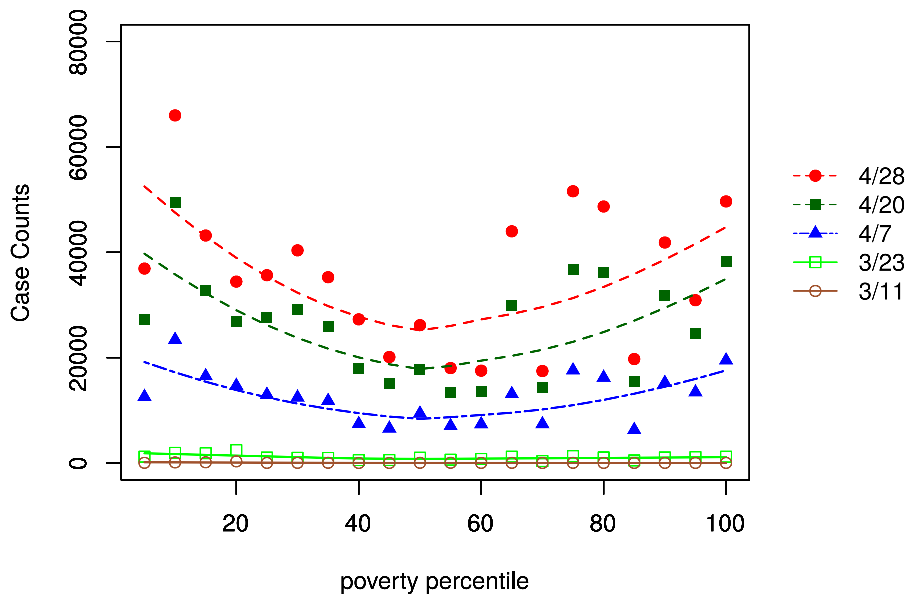

Figure C.1: **SARS-CoV-2 Confirmed Infections by Poverty Percentile Excluding New York**

Notes: The figure replicates Figure 4 except that we drop counties from the state of New York from the analysis. The source of data is USAFacts, as of April 28, 2020. We report the number of cumulative (infection) cases for March 11, March 23, April 7, April 20, and April 28 in 2020 for 20 county-groups (excluding counties from New York) ranked by poverty rate percentile. The curves are fitted using a smoothing method based on local linear regressions as described in Appendix A.

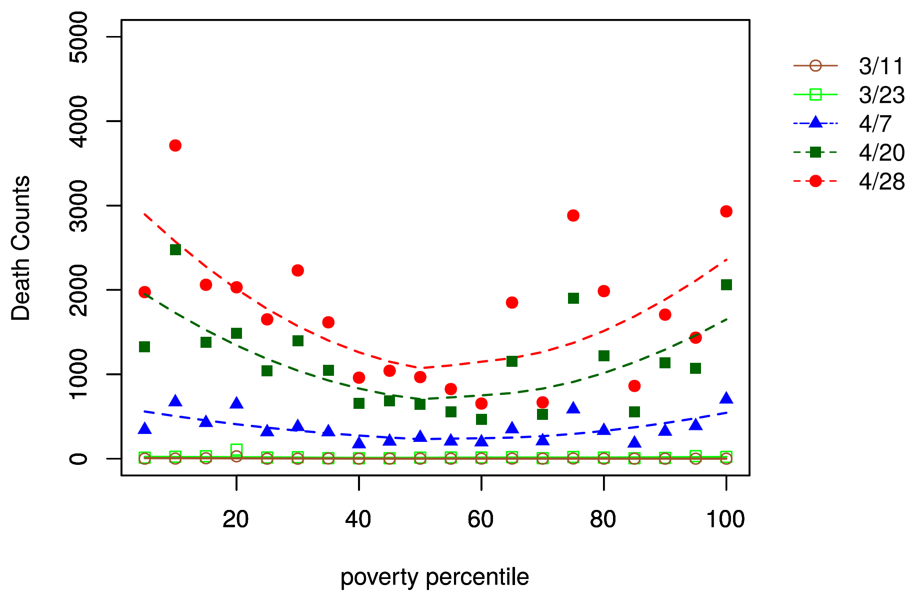

Figure C.2: **SARS-CoV-2 Confirmed Deaths by Poverty Percentile Excluding New York**

Notes: The figure replicates Figure 5 except that we drop counties from the state of New York from the analysis. The source of data is USAFacts, as of April 28, 2020. We report the number of cumulative deaths due to COVID-19 for March 11, March 23, April 7, April 20, and April 28 in 2020 for 25 county-groups (excluding counties from New York) ranked by poverty rate percentile. The curves are fitted using a smoothing method based on local linear regressions as described in Appendix A.

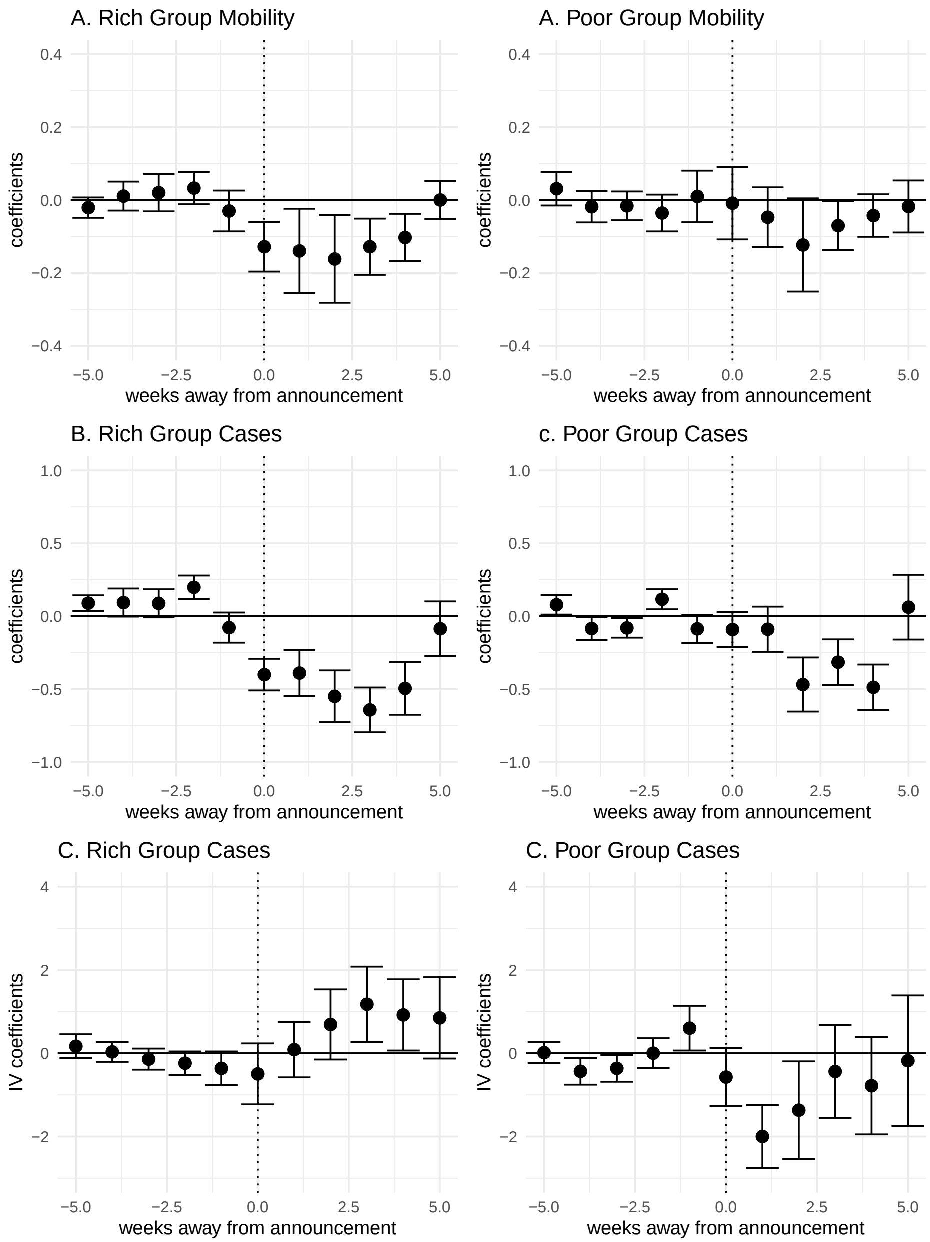

Figure C.3: **The Effects of Weather Shocks and Mandate on Mobility and COVID-19**

Notes: The figure is structured in a similar way as in Figure 9 but excludes New York from the analysis. **Panel A** uses log of mobility as the dependent variable, whereas **Panels B** and **C** use the log of weekly new cases. Panel A plots the coefficients on the interaction term between the county poverty group, weeks away from the mandate, and weather shock indicators as depicted in specification 1, where the figures on the left and right plot the estimates on κ_j and λ_j, respectively. Panel B plots the reduced form effect of stay-at-home mandate coupled with the extreme weather shock by plotting the estimates of κ_j and λ_j after estimating equation 2 on the left and right side of Panel B, respectively. Panel C shows the IV results for rich (left) and poor (right) county groups by plotting the estimates of κ_j and λ_j in specification 3.